Robust Classification of Subcellular Location Patterns in High-Resolution 3D Fluorescence Images

Chen, Xiang*, Velliste, Meel, Murphy, Robert F.
Departments of Biological Sciences and Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

In order to be fully functional, proteins need to be sorted to their correct destinations (organelles or other special subcellular compartments, such as the cytoskeleton). Knowledge of the subcellular location of a protein is valuable for predicting its function and for a complete understanding of its role in various cellular mechanisms. Fluorescence microscopy is frequently used to examine the subcellular distributions of proteins. However, the distributions are largely analyzed by visual examination, which is crude, labor-intensive and subjective. It also does not enable sensitive comparison of the patterns of proteins on a proteome-wide basis. Previously our group has designed sets of subcellular location features to objectively describe fluorescence images (both 2D and 3D) and achieved significant progress in distinguishing all major subcellular location patterns using these features. With developments in microscope technology over the past decade, 3D imaging is increasingly popular, and it is expected that a 3D image contains greater information content than a 2D image. Our initial feature set for describing a 3D image was solely based on morphological image processing. When applied to a dataset of 3D images of HeLa cells, this set of features achieved a classification accuracy of 91% on 10 subcellular location patterns using a back-propagation neural network. To improve the classification, we implemented two more sets of features, namely edge features and 3D Haralick texture features. The classification accuracy improved to 98% using a subset of the morphological, edge features, and 3D Haralick texture features chosen by stepwise discriminant analysis (a feature selection method). This represents a significant improvement in classification performance over the previous 3D feature set. The system is not only able to distinguish major subcellular patterns, but it can discriminate patterns of different Golgi proteins that are not distinguishable by humans. Overall, the results demonstrate that a fully automated system can be used to classify the subcellular patterns in 3D images using features that are relatively inexpensive to calculate. When applied to random protein tagging projects, the current method can objectively ascribe subcellular location patterns to tagged proteins, increasing throughput speed, reducing manpower and minimizing error rates. These features also enable high resolution comparison of subcellular location patterns and therefore provide a starting point for the clustering of different location patterns based on their similarity. The clustering approach would lead us to a better understanding of subcellular location patterns and the relationships among them, which is not only an inherent component in proteomics, but also desirable and necessary in systematically studying protein sorting mechanisms (e.g., by providing high resolution training data for systems aimed at predicting location from sequence). Lastly, the automated methods described here can be used to systematically determine how protein location changes in the presence of drugs, during development, and during disease processes.